

Ефимов А.А.

Alexander A. Yefimov

ДОСТУПНЫЙ СУПЕРКОМПЬЮТЕР В АКАДЕМИЧЕСКОЙ СРЕДЕ, ИЛИ
ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ В УЧЕБНЫХ АУДИТОРИЯХ.

AFFORDABLE SUPERCOMPUTER IN AN ACADEMIC ENVIRONMENT.
CLOUD COMPUTING IN CLASSROOMS

yefimov@mail.ru

ГОУ ВПО Уральский государственный педагогический университет
г. Екатеринбург

Технологии облачной обработки данных широко используются крупными коммерческими организациями, но в настоящее время они становятся доступны даже исследовательским лабораториям вузов или НИИ. В статье автор предлагает модель создания доступного кластера для удовлетворения потребностей академической среды в вычислительных мощностях.

Суперкомпьютеры, кластеры, облачная обработка данных, распределенные вычисления.

Cloud computing technologies are widely used by large corporations, but nowadays they become more available to research institutions. In this article author suggests a model of creating an affordable cluster for the academic environment needs.

Supercomputers, clusters, cloud computing, distributed computing.

История. Зарождение рынка.

На протяжении многих лет определение понятия «суперкомпьютер» не раз становилось предметом многочисленных споров и дискуссий.

Чаще всего авторство термина приписывается Джорджу Мишелю и Сиднею Фернбачу [5], которые в конце 60-х годов XX века работали в Ливерморской национальной лаборатории и компании Control Data Corporation. Тем не менее, известен тот факт, что ещё в 1920 [2. С. 4] году газета New York World рассказывала о «супервычислениях», выполняемых при помощи табулятора IBM, собранного по заказу Колумбийского университета.

В IT лексикон, термин «суперкомпьютер» вошёл благодаря распространению компьютерных систем Сеймура Крея, таких как, Control Data 6600, Control Data 7600, Cray-1, Cray-2, Cray-3 и Cray-4. Сеймур Крей разрабатывал вычислительные машины, которые становились основными вычислительными средствами правительственных, промышленных и академических научно-технических проектов США с середины 60-х до середины 90-х годов. Не случайно в то время одним из популярных определений суперкомпьютера было следующее: «любой компьютер, который создал Сеймур Крей». Сам Крей никогда не называл свои детища суперкомпьютерами, предпочитая использовать вместо этого обычное название «компьютер».

80-е годы прошлого века охарактеризовались появлением множества небольших конкурирующих компаний, занимающихся созданием высокопроизводительных компьютеров, однако к середине 90-х большинство из них оставили эту сферу деятельности, что даже заставило обозревателей заговорить о «крахе

рынка суперкомпьютеров». На сегодняшний день суперкомпьютеры являются уникальными системами, создаваемыми «традиционными» игроками компьютерного рынка, такими как IBM, Hewlett-Packard, NEC и другими, которые приобрели множество мелких, ранее существовавших компаний, вместе с их опытом и технологиями. Компания Cray Inc. по-прежнему занимает достойное место в ряду производителей суперкомпьютерной техники.

Из-за большой гибкости самого термина до сих пор распространены довольно нечёткие представления о понятии «суперкомпьютер». Шутливая классификация Белла и Нельсона, предлагала считать суперкомпьютером любой компьютер, весящий более тонны. Современные суперкомпьютеры действительно весят более тонны, однако далеко не каждый тяжёлый компьютер достоин чести считаться суперкомпьютером. В общем случае, суперкомпьютер – это компьютер значительно более мощный, чем доступные для большинства пользователей машины [7]. При этом скорость технического прогресса сегодня такова, что нынешний лидер легко может стать завтрашним аутсайдером.

Архитектура также не может считаться признаком принадлежности к классу суперкомпьютеров, чему есть много примеров. Так, многие суперкомпьютеры сейчас отличаются от «обычных» лишь количественно (больше процессоров, памяти) но не качественно, а в прошлом различие могло и вовсе сводиться к более быстрому блоку арифметико-логических операций, или даже менее существенным изменениям, дающим вычислительное превосходство на один порядок.

Большинство суперкомпьютеров 70-х оснащались векторными процессорами, а к началу и середине 80-х небольшое число (от 4 до 16) параллельно работающих векторных процессоров практически стало стандартным суперкомпьютерным решением. Конец 80-х и начало 90-х годов охарактеризовались сменой магистрального направления развития суперкомпьютеров от векторно-конвейерной обработки к большому и сверхбольшому числу параллельно соединённых скалярных процессоров.

Массивно-параллельные системы стали объединять в себе сотни и даже тысячи отдельных процессорных элементов, причём ими могли служить не только специальные, но и доступные в свободной продаже процессоры. Большинство массивно-параллельных компьютеров создавалось на основе процессоров с архитектурой RISC.

Наши дни, перерождение рынка

В конце 90-х годов высокая стоимость специализированных суперкомпьютерных решений и нарастающая потребность разных слоёв общества в доступных вычислительных ресурсах привели к широкому распространению компьютерных кластеров. Эти системы характеризует использование отдельных узлов на основе дешёвых и доступных комплектующих для серверов, объединённых в один вычислитель. Несмотря на кажущуюся простоту, кластеры довольно быстро заняли достаточно большой сегмент рынка, обеспечивая высочайшую производительность при минимальной стоимости.

В настоящее время суперкомпьютерами принято называть компьютеры с огромной вычислительной мощностью. Такие машины используются для работы

с задачами, требующими интенсивных вычислений. Примерами таких задач может послужить, прогнозирование погодно-климатических условий, моделирование ядерных испытаний и пр.

В этом кроется ещё одно отличие суперкомпьютеров от серверов и мэйнфреймов – компьютеров с высокой общей производительностью, призванных решать типовые задачи, такие как обслуживание больших баз данных или одновременная работа с множеством пользователей.

Иногда суперкомпьютеры используются для работы с одним-единственным приложением, использующим всю память и все процессоры системы; в других случаях они обеспечивают выполнение большого числа разнообразных приложений, решающих задачи от предсказания погоды или моделирования ядерных взрывов до анализа белковых структур и просчёта анимационных фильмов.

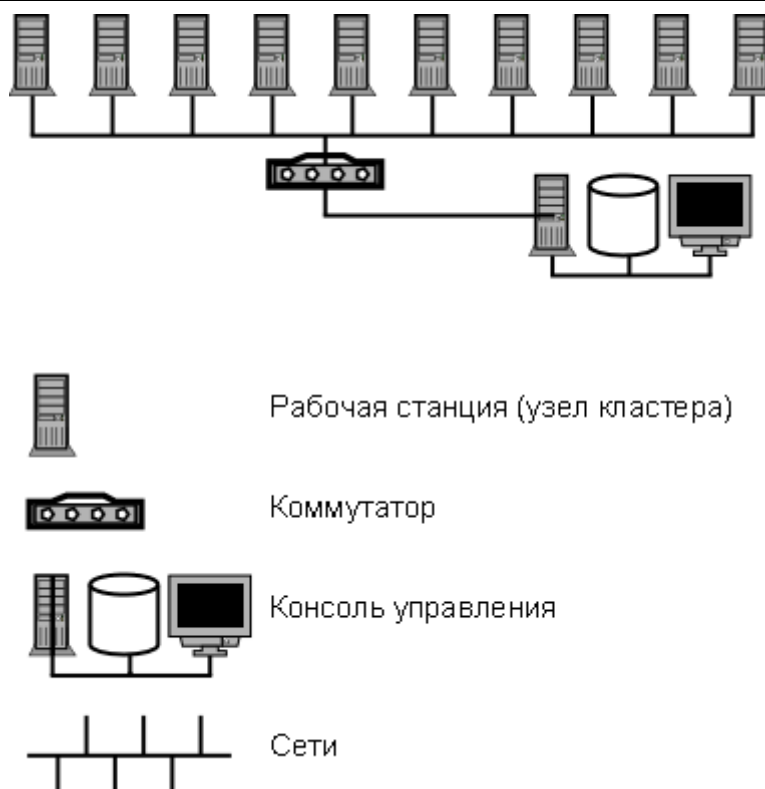
Cloud computing. Завтрашний день.

Новым словом в распределённых вычислениях стала Облачная обработка данных (Cloud Computing). Суть метода заключается в том, что программное обеспечение предоставляется пользователю как Интернет-сервис. Пользователь имеет доступ к собственным данным, но сложная инфраструктура, операционная система и собственно программное обеспечение, с которым он работает, скрыты за простым интерфейсом Интернет-браузера. "Облаком" метафорически называют Интернет, который скрывает все технические детали. Согласно определению, предложенному Б. Хайзом, [6], **Облачная обработка данных – это парадигма, в рамках которой информация постоянно хранится на серверах в сети Интернет и временно кэшируется на клиентской стороне, например на персональных компьютерах, игровых приставках, ноутбуках, смартфонах и т. д.**

Облачная обработка данных как концепция включает в себя понятия Программное обеспечение как услуга (Software as a Service, SaaS).

Удовлетворение потребности академической среды в вычислительных мощностях. Анализ возможностей.

Анализ вычислительных задач в академической среде показывает, что среди них есть и не типовые. Примеров таких задач великое множество. Это и моделирование процессов для нужд естественных или социальных наук, и распознавание текстов при оцифровке библиотечных фондов и многие другие задачи. Про прикладные задачи, стоящие на пути организации обучения курсам трёхмерного моделирования, или видеомонтажа речи не идёт, они и так на виду.



В настоящее время суперкомпьютеры, построенные в рамках парадигмы кластеризации в академической среде не редкость. Такие системы, например, решают задачи разработки методов моделирования для больших задач геофизики или создания информационно-вычислительных комплексов для решения прикладных задач математической физики в Уральском отделении Российской академии наук.

Перед научными институтами стоят исследовательские задачи, штат таких организаций обычно невелик, парк пользовательских компьютеров тоже. Классические кластерные суперкомпьютеры – одно из немногих решений для задач, требующих больших вычислений.

В учебных заведениях совсем другая ситуация. Редко какое подразделение вуза не имеет современной компьютерной аудитории, а в рамках института в целом парк компьютеров только в учебных аудиториях может составлять порядка тысяч штук.

Для построения простого кластера необходимы:

- компьютеры, которые будут разделять между собой вычислительную нагрузку или объединять свою вычислительную мощность,
- сеть, которая связывает эти компьютеры
- «центральный узел», который будет ставить задачи и выполнять функции сетевого хранилища.

Обычно, в сети любого учебного заведения уже имеются все эти компоненты, более того, вычислительная мощность компьютеров в учебных аудиториях используется крайне не оптимально, т.к. редко какие учебные задачи способны загрузить современный процессор, такой как Intel Core, обладающий производительностью порядка сотни миллионов операций с плавающей запятой в секунду.

Мы намеренно не затрагивали одну из важнейших составляющих кластера, но сейчас самое время рассмотреть её подробно. Речь идёт о специализированном программном обеспечении, распределяющем вычислительную нагрузку между элементами кластера и соединяющем результаты в единое целое. Рынок такого программного обеспечения бурно развивается и сейчас можно без труда назвать дюжину продуктов, таких как MOSIX, openMOSIX, MPICH, Microsoft WCCS, Intel Cluster Tools, способных собрать парк машин в учебных аудиториях в мощный суперкомпьютер, пригодный для решения широкого круга задач.

Правда, каждый компьютер необходимо настраивать соответствующим образом, устанавливать специализированные программы, а возможно даже модифицировать, или вовсе заменять операционную систему, что в ключе продолжения использования аудитории в качестве учебной, выглядит неприемлемым. Именно эту проблему и призвана решить облачная обработка данных.

Облачные вычисления. Ключ к прозрачности и масштабируемости

В качестве реализации рассмотрим следующий пример. Практически все современные браузеры обладают возможностью довольно быстро исполнять скриптовые языки, например JavaScript.

Благодаря этому появилась возможность использовать посетителей сайта как составные части кластера для обчёта больших массивов данных.

Естественно, алгоритм должен быть хорошо распараллеливаемым и отдельные части вычисления должны проходить быстро.

Примером таких вычислений может быть нахождение определителя матрицы или подсчет контрольных сумм.

Так же, поскольку процесс распределённых вычислений происходит без централизованного контроля оператором, необходимо проверять результат вычисления. Одним из способов может быть сравнение нескольких результатов от разных узлов с целью определить правильное методом "голосования".

Примером может послужить скрипт, который предоставил К. Тумалевич.

После инициализации скрипт получает с сервера список задач, в данном случае массив со строками. После чего устанавливает интервал на то, чтобы раз в 500 миллисекунд обработать одну задачу и отправить её на сервер.

После того, как список опустеет, все начинается сначала. Результат действия можно наблюдать в таблице в переменных `jobs`, `answers` и `this_stat`. Данный пример иллюстрирует, как в определенных случаях пользователей Интернет-приложения можно использовать как кластер в рамках облачной обработки данных.

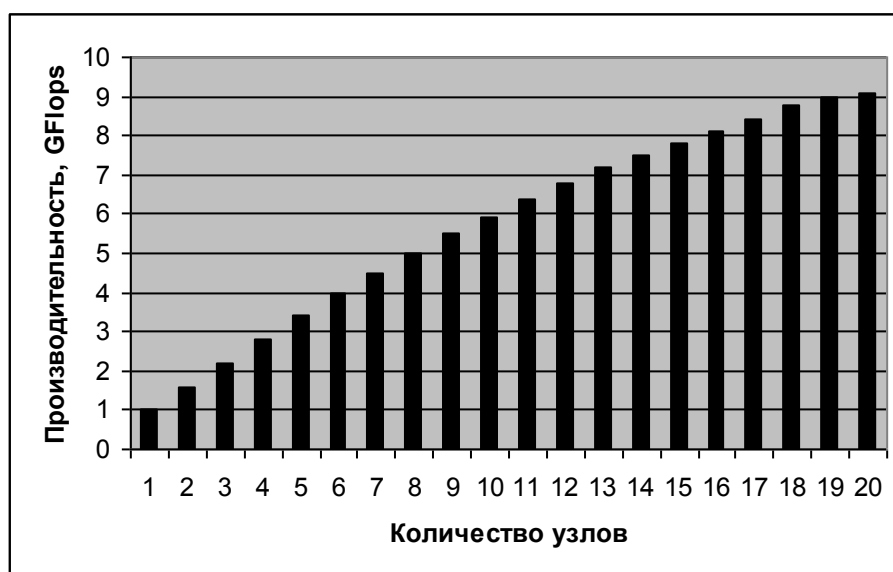
В сообществе Интернет-разработчиков такой способ кластеризации получил название «паразитные вычисления», но этический аспект этой проблемы (в ключе использования вычислительной мощности посетителей Интернет-ресурса без их согласия) выходит за рамки данной статьи. Сейчас важно лишь показать, что посредством специального Интернет-приложения, апплета или скрипта, можно моментально организовать высокопроизводительный кластер в короткие сроки. При этом узлы кластера не будут нуждаться в какой-либо дополнительной настройке

или дополнительном программном обеспечении и это решение будет легко масштабироваться.

Использование скриптовых языков программирования позволяет легко решать хорошо распараллеливаемые «математические задачи» под которые уже разработаны алгоритмы, но практически сводит на нет возможность решения задач распознавания текста. В настоящее время нами реализуется система распределённых вычислений на базе библиотеки MPI (Message Passing Interface – интерфейс передачи сообщений) и технологии ActiveX для построения масштабируемого кластера, не требующего настройки вычислительных узлов, и предоставляющего привычный интерфейс доступа для параллельных вычислений. Библиотека MPI является наиболее распространённым стандартом интерфейса обмена данными в параллельном программировании. Существуют его реализации для большого числа компьютерных платформ. Технология ActiveX, в свою очередь, может быть без проблем использована в браузерах семейства Internet Explorer.

Теперь о том, что касается сухих цифр. В некоторых задачах кластеризованная учебная аудитория из двадцати компьютеров с процессорами Intel Core 2 DUO E8400 и двумя гигабайтами оперативной памяти, подключенных к коммутатору Trendnet TEG-240WS (Gigabit Enhernet, 24 порта) показывает практически линейный рост производительности от одного до десяти GFlops при использовании от двух до двадцати узлов в кластера.

Зависимость роста производительности от роста числа узлов кластера проиллюстрирована на диаграмме 1.



В правой части диаграммы видно, что характер зависимости изменяется, рост производительности есть, но скорость его падает. Это как связано с тем, что наш программный прототип находится лишь на начальном этапе разработки и практически не оптимизирован, так и с дефицитом аппаратных ресурсов маршрутизатора. Дальнейшее увеличение узлов кластера не приносит устойчивого роста производительности.

Для сравнения была исследована зависимость роста вычислительной мощности «традиционного» кластера. Аудитория осталась прежней, но в качестве

кластеризующего программного обеспечения был использован пакет Intel Cluster Tools.

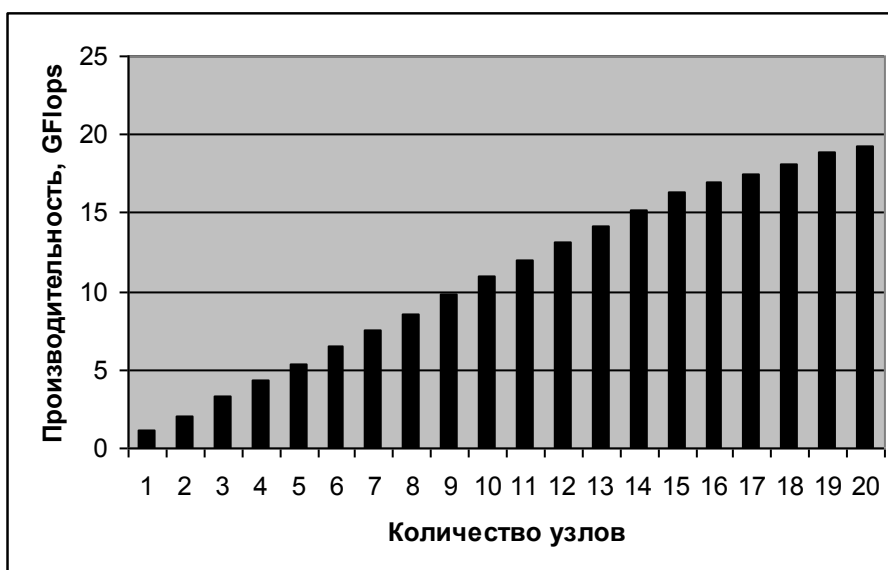


Диаграмма 2 говорит сама за себя. Традиционный кластер демонстрирует как лучшую производительность так и масштабируемость.

Здесь важно отметить, что для получения результатов, на «традиционном» кластере потребовалось установить на компьютеры специальную операционную систему (нами использовался Debian GNU/Linux 5), специального программного обеспечения (Intel Cluster Tools) и дополнительной настройки. В рамках «облачного» кластера удалось получить гораздо меньший результат, но во время тестов узлы кластера использовались как учебные компьютеры. На них был лишь запущен WEB-браузер с низким приоритетом.

Идея «облачного» кластера в частности и паразитических вычислений в целом относительно нова, но в наши дни проблема эффективного использования ресурсов как никогда актуальна.

Технологии облачной обработки данных широко используются крупными коммерческими организациями, но в настоящее время они становятся доступны даже исследовательским лабораториям вузов или НИИ и компаниям среднего уровня. Благодаря использованию недорогого потребительского ПО и использованию продуктов с открытым исходным кодом, проведение экспериментов с помощью этих технологий довольно просто, даже при очень маленьком бюджете. Тем не менее, использование технологии облачной обработки данных требует определённой квалификации и опыта. Поэтому ее изучение должно занять достойное место в учебных планах, как технических специальностей, так и технологических специализаций.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК:

1. Тумалевич К. Использование процессорных мощностей посетителей сайта. Паразитические вычисления [Электронный ресурс]. - Режим доступа: <http://tumalevich.pp.ru/2009/3/4/307>(дата обращения 10.03.2009)
2. Axner L. High Performance Computational Hemodynamics with the Lattice Boltzmann Method. - Amsterdam, 2007. - P. 4

3. Bell G. The future of high performance computers in science and engineering / Gordon Bell // Communications of the ACM. - 1989. Vol. 32, N 9. - P. 1091-1101
4. Duff I. S. The use of supercomputers in Europe / Iain S. Duff // Computer Physics Communications. - 1985. - N 37. - P. 15-25
5. Fernbach S. Supercomputers – past, present, prospects / Sidney Fernbach // Future Generation Computer Systems. - 1984. - Vol. 1, Issue 1. - P. 23-38
6. Hayes B. Cloud computing / Brian Hayes // Communications of the ACM. - 2008. - Vol. 51, Issue 7. - P. 9-11
7. IEEE Standard for Information Technology-POSIX@-Based Supercomputing Application Environment Profile [Approved June 14,1995. - IEEE Std 1003.10-1995] [Электронный ресурс] / Portable Applications Standards Committee of the IEEE Computer Society. - Режим доступа: http://ieeexplore.ieee.org/xpls/abs_all.jsp?tp=&isnumber=10080&arnumber=478420&punumber=3414 (дата обращения 23.03.2009)
8. Oyanagi Y. Future of supercomputing/ Yoshio Oyanagi // Journal of Computational and Applied Mathematics. - 2002. - N 149. - P. 147–15

Зайцева Н.А., Ушкова В.И., Филиппов В.В.

Zayceva N.A., Ushkova V.I., Filippov V.V.

ИНТЕРНЕТ-ТЕСТИРОВАНИЕ ПО ФИЗИКЕ

INTERNET-TESTING OF PHYSICS

natalzay@yandex.ru

ЕБАКУ

г. Екатеринбург

Проведен анализ Интернет-тестирования как элемента управления качеством образования на примере дисциплины «физика», сформулированы замечания и предложения по оптимизации кодификатора, содержания и оценки результатов тестирования.

The analysis of Internet testing as element of quality management of education on an example of physics is carried out. Remarks and offers on optimization of codificator, contents and results estimation of testing are formulated.

Интернет-тестирование проводится Национальным аккредитационным агентством в сфере образования с 2005 г. в рамках эксперимента по введению Федерального экзамена в высшем профессиональном образовании (ФЭПО). Содержанием эксперимента является проведение компьютерного Интернет-тестирования в части внешней оценки уровня подготовки студентов на соответствие требованиям государственных образовательных стандартов (ГОС). 22 июня 2009 года закончился девятый этап Интернет-экзамена. В тестировании приняло участие 1299 вузов и филиалов вузов, 557 ссузов и филиалов ссузов из 82 регионов РФ, а также из стран СНГ. Всего было получено 1394534 результатов тестирования.

В летнюю сессию 2008-2009 учебного года наш военный институт впервые прошел процедуру Интернет-тестирования (специальность 140601.65-